Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2005 Proceedings

Americas Conference on Information Systems (AMCIS)

2005

Studying the Structure of Terrorist Networks: A Web Structural Mining Approach

Jialun Qin
The University of Arizona, qin@eller.arizona.edu

Jennifer J. Xu
The University of Arizona, jxu@eller.arizona.edu

Yihu Zhou
The University of Arizona, yiluz@eller.arizona.edu

Hsinchun Chen
The University of Arizona, hchen@eller.arizona.edu

Studying the Structure of Terrorist Networks: A Web Structural Mining Approach

Jialun Qin
Artificial Intelligence Lab, Department of Management Information Systems, The University of Arizona
qin@eller.arizona.edu

Jennifer J. Xu
Artificial Intelligence Lab, Department of Management Information Systems, The University of Arizona
jxu@eller.arizona.edu

Yilu Zhou
Artificial Intelligence Lab, Department of Management Information Systems, The University of Arizona
yiluz@eller.arizona.edu

Hsinchun Chen
Artificial Intelligence Lab, Department of Management Information Systems, The University of Arizona
hchen@eller.arizona.edu

ABSTRACT

Because terrorist organizations often operate in network forms where individual terrorists collaborate with each other to carry out attacks, we could gain valuable knowledge about the terrorist organizations by studying structural properties of such terrorist networks. However, previous studies of terrorist network structure have generated little actionable results. This is due to the difficulty in collecting and accessing reliable data and the lack of advanced network analysis methodologies in the field. To address these problems, we introduced the Web structural mining technique into the terrorist network analysis field which, to the best our knowledge, has never been done before. We employed the proposed technique on a Global Salafi Jihad network dataset collected through a large scale empirical study. Results from our analysis not only provide insights for terrorism research community but also support decision making in law-reinforcement, intelligence, and security domains to make our nation safer.

Keywords

Terrorism, social network analysis, web structural mining, visualization.

INTRODUCTION

Terrorism threats span personal, organizational, and societal levels and have far-reaching economic, psychological, political, and social consequences. Only with a thorough understanding of how the terrorist organizations function can we defend against these threats. Previous studies showed that terrorist organizations often operate in network forms where individual terrorists collaborate with each other to carry out attacks (Krebs, 2001). Thus, we could gain valuable knowledge about the terrorist organizations by studying various structural properties of terrorist networks. Such knowledge may help authorities develop efficient and effective disruptive strategies and measures.

Terrorism-related research domain has experienced tremendous growth since September 11th; however, studies of terrorist networks have generated little actionable results. This is mainly due to the difficulty in collecting and accessing reliable data and the lack of advanced network analysis methodologies in this field. To address these problems, we report in this paper a case study of the analysis of the structure of a very large global terrorist network, the Global Salafi Jihad (GSJ) network using methods and techniques from several relevant areas such as Web structural mining and social network analysis. We consider our case study unique and beneficial from three different perspectives. First, unlike most previous studies which used unreliable data sources such as news stories and media-generated incident databases, our study was based on reliable data collected in a large-scale in-depth empirical study on the GSJ network (Sageman, 2004). Second, our study introduced multiple advanced network analysis methodologies into the study of terrorist networks including the Web structural mining techniques which, to the best our knowledge, has never been used in this domain. Third, our results not only provide insights for terrorism research community but also support the decision-making within law-reinforcement, intelligence, and security domains to make our nation safer.

The remainder of the paper is organized as follows. Section 2 reviews various network analysis studies in different domains in relation to terrorist network analysis. In section 3, we provide some background information on the GSJ network and briefly describe how the GSJ network dataset was collected through the empirical study. In section 4, we present our methodologies and report our findings from the analysis. Section 5 concludes this paper with implications and future directions.

RELATED WORKS

In this section, we review a few network analysis methodologies widely employed in other domains: social network analysis and Web link structure analysis. These techniques can be used to analyze terrorist networks. Different techniques reveal different perspectives of terrorist networks.

Social Network Analysis

Social network analysis (SNA) is used in sociology research to analyze patterns of relationships and interactions between social actors in order to discover an underlying social structure (Scott, 1991, 2001; Wasserman and Faust, 1994). A number of quantitative SNA methods have been employed to study organizational behavior, inter-organizational relations, citation analysis, computer mediated communication, and many other domains (Galaskiewicz and Krohn, 1984; Garton, Haythornthwaite and Wellman, 1999; Kleinberg, 1998). SNA has recently been recognized as a promising technology for studying criminal organizations and enterprises (McAndrew, 1999; Sparrow, 1991). Studies involving evidence mapping in fraud and conspiracy cases have recently been added to this list (Baker and Faulkner, 1993; Saether and Canter, 2001).

In SNA studies, a network is usually represented as a graph, which contains a number of nodes (network members) connected by links (relationships). SNA can be used to identify key members and interaction pattern between sub-groups in terrorist networks. Several centrality measures can be used to identify key members who play important roles in a network. Freeman (Freeman, 1979) provided definitions of the three most popular centrality measures: degree, betweenness, and closeness.

Degree measures how active a particular node is. It is defined as the number of direct links a node *a* has:

$$C_D(a) = \sum_{i=1}^{n} c(i,a)$$

where n is the total number of nodes in a network, $c(i, a)$ is a binary variable indicating whether a link exists between nodes *i* and *a*. A network member with a high degree could be the leader or "hub" in a network.

Betweenness measures the extent to which a particular node lies between other nodes in a network. The betweenness of a node *a* is defined as the number of geodesics (shortest paths between two nodes) passing through it:

$$C_B(a) = \sum_{i<}^{n}\sum_{j}^{n} g_{ij}(a)$$

where $g_{ij}(a)$ indicates whether the shortest path between two other nodes *i* and *j* passes through node *a*. A member with high betweenness may act as a gatekeeper or "broker" in a network for smooth communication or flow of goods (e.g., drugs).

Closeness is the sum of the length of geodesics between a particular node *a* and all the other nodes in a network. It actually measures how far away one node is from other nodes and sometimes is called "farness" (Baker and Faulkner, 1993; Freeman, 1979, 2000):

$$C_C(a) = \sum_{i=1}^{n} l(i,a)$$

where $l(i,a)$ is the length of the shortest path connecting nodes *i* and *a*.

Web Structural Analysis

The Web is one of the largest and most complicated networks in the world. The Web, as a network of Web pages connected by hyperlinks, bears some similarities with social networks because previous studies have shown that the link structure of the Web represents a considerable amount of latent human annotation (Gibson, Kleinberg and Raghavan, 1998). For example, when there is a direct link from page A to page B, it often means that the author of page A recommends page B because of its relevant contents. Moreover, similarly to citation analysis in which frequently cited articles are considered to be more important, Web pages with more incoming links are often considered to be better than those with fewer incoming links. Co-citation is another concept borrowed from the citation analysis field that has been used in link-based analysis algorithms. Web pages are co-cited when they are linked to by the same set of parent Web pages and heavily co-cited pages are often relevant to the same topic. Co-citation is particularly helpful in finding relevant pages in some domains where pages with similar contents avoid linking to each other (e.g., commercial domains where providers of similar online contents are competitors). Researchers have developed many algorithms to judge the importance and quality of Web pages using the criteria mentioned above. PageRank is one of the most popular algorithms.

The PageRank algorithm is computed by weighting each incoming-link to a page proportionally to the quality of the page containing that incoming-link (Cho, Garcia-Molina and Page, 1998). The quality of these referring pages is also determined by PageRank. Thus, the PageRank of a page p is calculated recursively as follows:

$$\text{PageRank(p)} = \ 1 - d + d \times \sum_{\text{all q links to p}} \frac{PageRank\ (q)}{c(q)}$$

where d is a damping factor between 0 and 1 and c(q) is the number of out-going links in q.

PageRank is originally designed to calculate the importance of Web pages based on the Web link structure and is used in the commercial search engine Google (Brin and Page, 1998) to rank the search results. However, it can also be used to determine the importance of social actors in a proper social network where links imply similar “recommendation” or “endorsement” relationships as the hyperlinks in Web graph. In a co-authorship network, a link between authors implies the mutual endorsement relationship between them and the PageRank algorithm can be used to rank the authors based their importance in this co-authorship network. In the co-authorship analysis study conducted by Liu et al. (2004), PageRank was used as one of the author ranking criteria along with other traditional SNA centrality measures. Similarly, we believe that PageRank can also used to rank the importance of terrorists within a properly constructed terrorist network.

GLOBAL SALAFI JIHAD NETWORK

The Global Salafi Jihad (GSJ) is part of a violent worldwide terrorism movement. It is a new form of terrorism which threatens the worlds in different and horrifying ways from previous forms of this scourge. It mainly targets the West, but its reckless operations and indiscriminately slaughter masses of humanity of all races and religions. With Al Qaeda as its vanguard, the GSJ includes many terrorist groups with members from different countries and forms a large global terrorist network. Through this network, the GSJ have successfully planned and launched many large-scale attacks against civilians across different countries. Examples include the 9/11 tragedy in 2001, the bombing in Bali in 2002, and the bombing in Morocco in 2003.

Collecting data on the GSJ terrorist presents many challenges, mostly because of a general lack of information. The GSJ data we used in this study was collected through a long-term empirical study on the GSJ members. The sources of information we used to collect data from were all in the public domain. The information was often inconsistent. We considered the source of information in selection facts to include in the dataset. In decreasing degrees of reliability, the information sources we favored include transcripts of court proceedings involving GSJ terrorists and their organizations; followed by reports of court proceedings; then corroborated information from people with direct access to the information provided; uncorroborated statements from people with the access; and finally statements from people who had heard the information secondhand. Data collected from these multiple sources were cross-validated to ensure maximum accuracy.

The final dataset consists of the profile information of 366 GSJ terrorists roughly divided into 4 clumps based on their geographical origins: central member, core Arab, maghreb Arab, and Southeast Asian. The central member clump mainly consists of the key Al Qaeda members. They take the leading position in the whole GSJ network. The core Arab clump consists of GSJ terrorists from core Arabic countries such as Saudi Arabia and Egypt. The maghreb Arab clump consists of GSJ terrorists from North African countries such as Morocco and Algeria. Finally, the Southeast Asian clump consists of terrorists from Jemaah Islamiyah centered in Indonesia and Malaysia.

The data collected for each of the 366 terrorists includes a set of sociological features (e.g., geographical origins, original socio-economic status, education, occupation, etc) and individual psychological (e.g., mental illness, personality, pathological narcissism, etc) features that could be the explanations of why these people became terrorists. More importantly, the data also captures all known relationships and interactions between these 366 GSJ terrorists. These relationships and interactions include personal relationships (e.g., acquaintance, friend, relative, and family member), religious relationships (following the same religious leader), operational interactions (participating in the same attacks), and other relationships. The dataset is presented in a form a spreadsheet with each raw containing the basic features of a certain GSJ member as well as all the other members that are related to this member through the various relationships or interactions mentioned above. We then calculated the “distance” between each pair of terrorists in the network based on the number of relationships between them and visualized the network using multidimensional scaling (MDS) technique. Our visualization provides an intuitive and clear view of the overall GSJ network (See Figure 1).

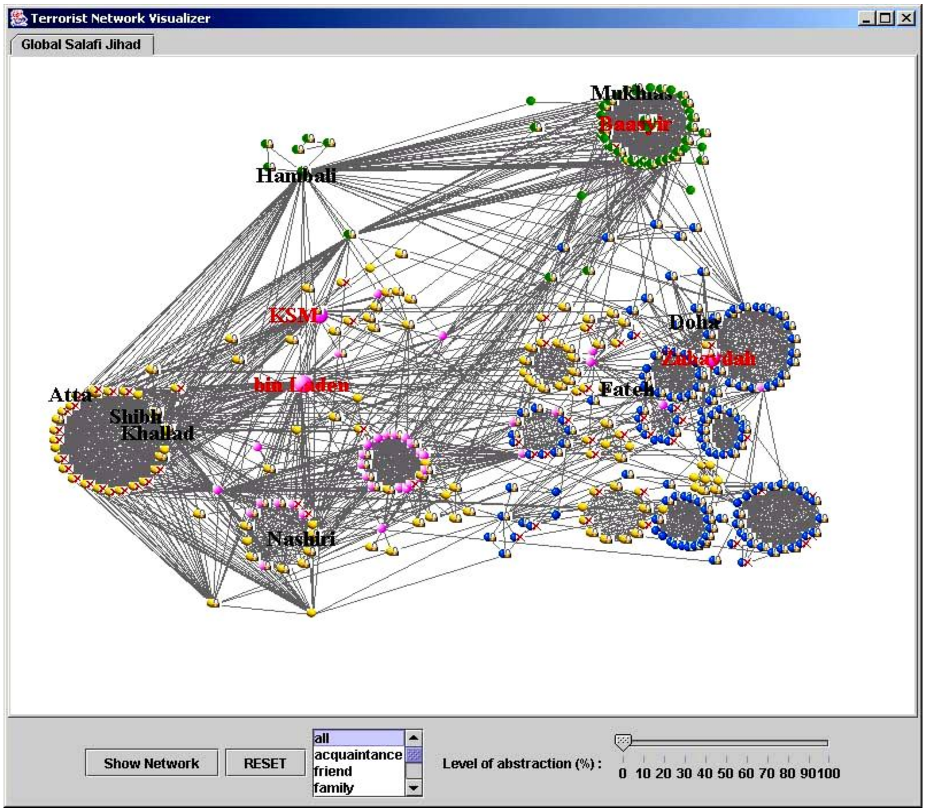

Figure 1. Visualization of the Full GSJ Network

Figure 1 is the visualization of the GSJ network with all types of relations. Each node represents a terrorist. A link represents a social relation. The four terrorist clumps are color-coded: red for central member clump, yellow for core Arab clump, blue for Maghreb Arab, and green for Southeast Asian clump.

NETWORK ANALYSIS ON THE GSJ DATASET

To better understand how the GSJ network works, we employed the proposed Web structural mining approach and the traditional SAN techniques on the GSJ network dataset. In this section, we describe our analysis procedures and report our findings.

Web Structural Mining Approach

Our assumption in applying the Web structural mining approach is that, in a social network, not all the members play equal roles. Instead, some members may have stronger social influences or higher social status than the others. In a terrorist network context, a terrorist may act as a leading role and pass directions and orders to a group of terrorists who have lower status than him and at the same time he is also receiving directions and orders from someone who has higher status. Such unequalized social relationships between the terrorists may hold special interest for experts to study the terrorist organization behavior. To study the different social status and relationships in a terrorist network context, we borrowed the link analysis methodology from the Web structural mining area and applied it on our GSJ network data.

The core link algorithm we employed was the PageRank algorithm because it was used in previous studies to calculate the “importance” of authors within an authorship network. The link analysis we conducted on the GSJ network is described as follows.

First, we used the PageRank algorithm reviewed in section 2 to calculate a "social importance" score for each of the terrorists in the network. In this process, the PageRank algorithm will rank a terrorist higher if 1) he links to more other members in the network and 2) he links to other members with high importance scores in the network. Similarly to the degree measure, high importance scores given by the PageRank algorithm are also indications of leading roles in the terrorist network. However, PageRank algorithm determines the importance of a specific member based on the structure of the whole network; while degree measure make the some judgment only based on very limited, local structural information.

After the importance scores for all the members in the GSJ network were calculated, for each member in the network, the neighboring member with the highest important scores was identified. The assumption here is that the most important neighboring member for a terrorist may well be the local leader that the terrorist directly report to. We then draw a directional link from each of the terrorists to their local leaders to visualize the terrorist social hierarchy and this graph is called a Authority Derivation Graph (ADG) (Toyoda and Kitsuregawa, 2001). Figure 2 shows the ADG of the GSJ network.

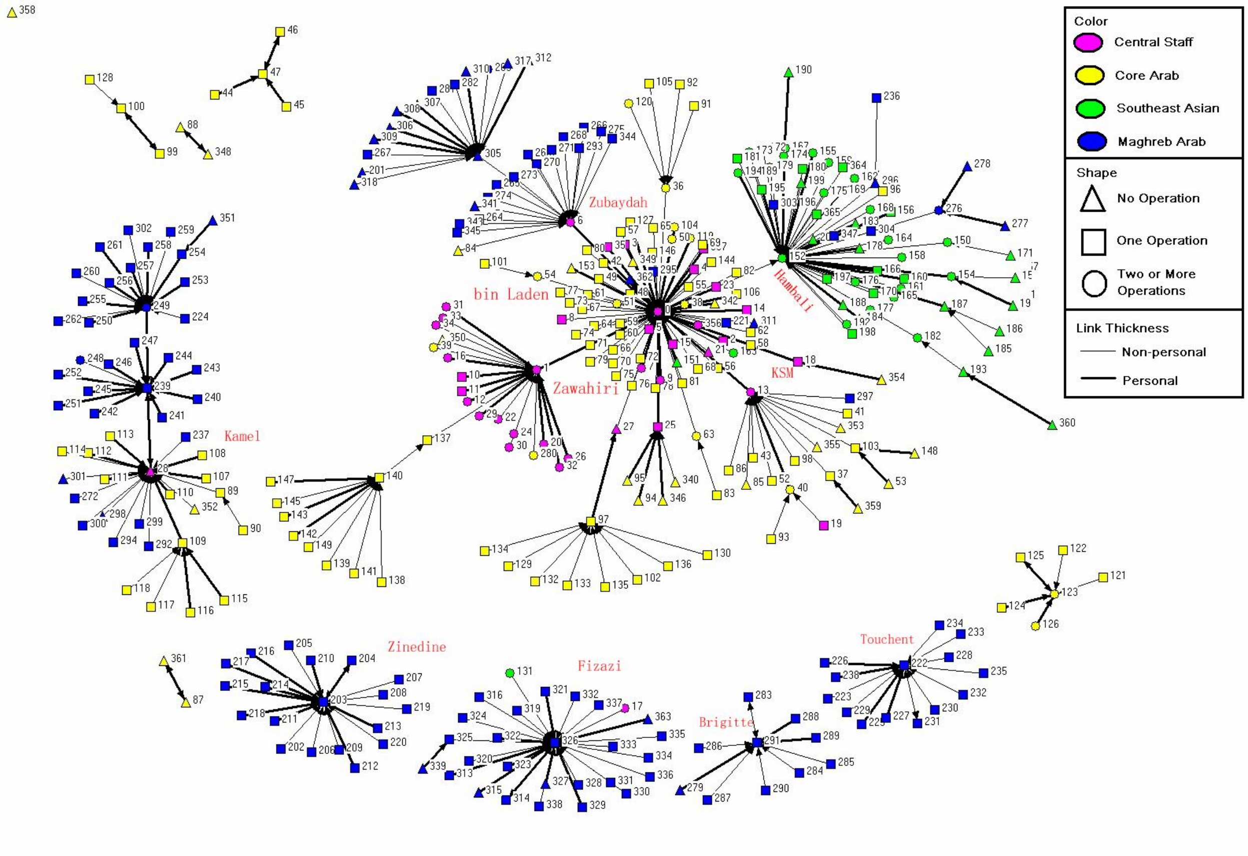

Figure 2. The Authority Derivation Graph of the GSJ Network

In the ADG, each node represents a terrorist in the GSJ network. A link pointing from terrorist *A* to terrorist *B* means that *B* has the highest rank among all members who have direct relationships with *A* and it is likely that, in their interaction, *B* acts as the role of "leader" and *A* acts as the role of "follower." The color of a node indicates which clump the member belongs to and the shape of a node indicates how many attacks the member has been involved in. The thickness of the links between nodes indicates the type of relationship between the members. A thick link means there are personal relationships (kinship, family, friends, acquaintance, etc) between two members while a thin link means there are only operational relationships (involved in the same attack) between two members.

The ADG of the GSJ terrorist network contains a large central component and several small and relatively autonomous components. 2. The central component, consisting of key Al Qaeda members, has a more traditional hierarchy or "corporate structure". We can clearly see that bin Laden has the highest status or, in other words, he is the leader of the whole GSJ network. Several major Lieutenants serve as the first level underlings and the middle-person between bin Laden and key members of the other 3 clumps. More specifically, Hambali is the middle-person between bin Laden and the Southeast Asian clump; Zubaydah serves as the middle-person between bin Laden and the Maghreb Arab clump; Zawahiri connects bin Laden to the remainder of the Central member clump; and KSM acts as the middle-person between bin Laden and the core Arab clump.

Except for the central component, the other components in the ADG have smaller size and shorter average shortest paths. This overall structure of the ADG suggests that the GSJ network may functions a “holding company” model, with Al Qaeda as the “umbrella organization” in charge of planning and many small independent groups as “operating divisions”. Such a model allows effective planning of attacks by having Al Qaeda as the “master brain” of the whole network and reduces the risk of being disrupted by leaving the operations to the smaller groups that have minimum interactions with the central members.

Another interesting observation we made from the ADG is the difference in the link types between different types of members in the network. We found that 65% of the links between the leaders (members with incoming links) are personal links (acquaintances, friends, relatives, and family members), while only 38% of the links between the leaders and the followers (members with no in-coming links) are personal links. Such differences in the link types between different members were also demonstrated in some other illegal networks such as drug dealer networks. The high percentage of personal relationships between the leaders forms the trustworthy “backbone” of the GSJ network and the low percentage of personal relationships between other members and the core members helps keep the network decentralized, covert, and less vulnerable.

Social Network Analysis Approach

For comparison purposes, we also conducted traditional centrality analysis on the GSJ dataset to identify the key members within the network.

For each terrorist, three centrality measures were calculated: degree, betweenness, and closeness. Degree measure was used to identify the leaders of each clump in the GSJ network. High degrees indicate high levels of activity and wide social influence, which means the members with high degrees are likely to be the leaders of their local networks. Gatekeepers, members with high betweenness, hold special interest for terrorist experts because gatekeepers are usually the contact person between several terrorist groups and play important roles in coordinating terrorist attacks. The closeness measure was used differently from the previous two centrality measures. Instead of terrorists with high closeness, we identified those with low closeness whom are usually called outliers in SNA literatures. Outliers are of special interest because previous literature showed that, in illegal networks, outliers could be the true leaders. They appear to be outliers because they often direct the whole network from behind the scene, which prevents authorities from getting enough intelligence on them. Table 1 summarizes the top 5 terrorists ranked by the 3 centrality measures in each of the 4 clumps.

Ranking	Leader (Degree)	Gatekeeper (Betweenness)	Outlier (Closeness)
Central Member			
1	Zawahiri	bin Laden	Khalifah
2	Makkawi	Zawahiri	SbinLaden
3	Islambuli	Khadr	Ghayth
4	bin Laden	Sirri	M Atef
5	Attar	Zubaydah	Sheikh Omar
Core Arab			
1	Khallad	Harithi	Elbaneh
2	Shibh	Nashiri	Khadr4
3	Jarrah	Khallad	Janjalani
4	Atta	Johani	Dahab
5	Mihdhar	ZaMihd	Mehdi
Southeast Asian			

1	Hambali	Baasyir	Siliwangi
2	Baasyir	Hambali	Fathi
3	Mukhlas	Gungun	Naharudin
4	Iqbal	Muhajir	Yunos2
5	Azahari	Setiono	Maidin
Maghreb Arab			
1	Doha	Yarkas	Mujati
2	Benyaich2	Zaoui	Parlin
3	Fateh	Chaib	Mahdjoub
4	Chaib	DavidC	Zinedine
5	Benyaich1	Maaroufi	Ziyad

Table 1. Terrorists with Top Centrality Ranks within Each Clump

After showing our ADG and SNA results to the domain experts, we confirmed that the overall structure of the ADG and key members identified by the centrality measures matched the experts' knowledge on the terrorism organization. Members with high degree measures are also known by the experts as the leaders of the clumps in real world. Also such members would appear to be the "hubs" in the ADG analysis results. For example, Osama bin Laden, the leader of the central member clump, had 72 links to other terrorists and ranked the second in degree and he appears to be the center member in the ADG of the whole GSJ network. Moreover, the experts mentioned that each clump has a Lieutenant who acts as an important connector between the clumps. For example, Zawahiri, Lieutenant of the central member clump, connects the central member clump and the core Arab clump together. Hambali, Lieutenant of the Southeast Asian clump, connects the Southeast Asian clump and the central member clump. These Lieutenants were also correctly identified by the SNA analysis for their high betweenness. Moreover, these Lieutenants appear to be the branch nodes on the ADG which connects the root nodes (leaders) and the leaf nodes (low-level members).

CONCLUSION

It is very important for us to understand the functions and structures of terrorist networks to win the battle against terror. In this study, we employed several advanced network analysis techniques on a GSJ network dataset collected through a large scale empirical study. We applied Web structural mining methodologies in the GSJ network analysis. This approach, to the best of our knowledge, has never been used in this domain before and it helps us study the terrorist organization structure under a social hierarchy assumption. This may provide insights into better understanding of terrorist organization behavior. We also applied traditional SNA analysis on the GSJ dataset. Domain experts confirmed that both our Web structural mining results and the SNA results matched their on knowledge and the Web structural mining analysis provides new insights to the terrorist network analysis domain.

We have several future research directions to pursue. First, we are working with terrorism experts to fine tuning our algorithms to generate more accurate results. Second, we plan to extend the scope of our project to other types of illegal networks such as crime networks. Third, we want to add time-series analysis to get a more comprehensive understanding of the evolution and dynamics of terrorism networks (Sageman, 2004).

ACKNOWLEDGMENTS

We would like to sincerely thank our domain expert, Dr. Marc Sageman from the University of Pennsylvania, for providing us the dataset and many helpful comments and suggestions. We would also like to thank other AI Lab members who have contribute to the tools we used in this study and provided various help, more specifically, Daning Hu, Guanpi Lai, and Cathy Larson.

REFERENCES

1. Baker, W. E. and Faulkner, R. R. (1993) The Social Organization of Conspiracy: Illegal Networks in the Heavy Electrical Equipment Industry, *American Sociological Review*, 58, 12, 837-860.
2. Brin, S. and Page, L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine, *Computer Networks and ISDN Systems*, 30, 1–7.
3. Cho, J., Garcia-Molina, H. and Page, L. (1998) Efficient Crawling through URL Ordering, *Proceedings of the 7th International WWW Conference*, Brisbane, Australia.
4. Freeman, L. C. (1979) Centrality in Social Networks: Conceptual Clarification, *Social Networks*, 1, 215-240.
5. Freeman, L. C. (2000) Visualizing Social Networks, *Journal of Social Structure*, 1, 1.
6. Galaskiewicz, J. and Krohn, K. (1984) Positions, roles, and dependencies in a community interorganization system, *Sociological Quarterly*, 25, 527-550.
7. Garton, L., Haythornthwaite, C. and Wellman, B. (1999) Studying online social networks, Doing Internet Research, S. Jones, Sage Press.
8. Gibson, D., Kleinberg, J. and Raghavan, P. (1998) Inferring Web Communities from Link Topology, *Proceedings of the 9th ACM Conference on Hypertext and Hypermedia*, Pittsburgh, Pennsylvania, USA.
9. Kleinberg, J. (1998) Authoritative Sources in a Hyperlinked Environment, *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, San Francisco, CA.
10. Krebs, V. E. (2001) Mapping networks of terrorist cells, *Connections*, 24, 3, 43-52.
11. Liu, X., Bollen, J., Nelson, M. L. and Van de Sompel, H. (2004) All in the Family? A Co-Authorship Analysis of JCDL Conferences (1994-2003), *Proceedings of the IEEE/ACM Joint Conference on Digital Libraries 2004*, Tucson, AZ.
12. McAndrew, D. (1999) The Structural Analysis of Criminal Networks, The Social Psychology of Crime: Groups, Teams, and Networks, *Offender Profiling Series*, III., Dartmouth, Aldershot.
13. Saether, M. and Canter, D. V. (2001) A structural analysis of fraud and armed robbery networks in Norway, *Proceedings of the 6th International Investigative Psychology Conference*, Liverpool, England.
14. Sageman, M. (2004) Understanding Terror Networks, Philadelphia, PA, University of Pennsylvania Press.
15. Scott, J. (1991) Social Network Analysis, London, Sage Press.
16. Scott, M. (2001) War's new front, *CIO Insight*, 82-83.
17. Sparrow, M. K. (1991) The Application of Network Analysis to Criminal Intelligence: An Assessment of the Prospects, *Social Networks*, 13,.251-274
18. Toyoda, M. and Kitsuregawa, M. (2001) Creating a Web Community Chart for Navigating Related Communities, *Proceedings of ACM Conference on Hypertext and Hypermedia*, Århus, Denmark.
19. Wasserman, S. and Faust, K. (1994) Social Network Analysis: Methods and Applications, Cambridge, Cambridge University Press.